

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Nomi deverbali in MIDIA

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1624728> since 2017-02-14T12:30:09Z

Publisher:

Franco Cesati

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

LIVIO GAETA*

NOMI DEVERBALI IN MIDIA**

1. Introduzione

Com'è noto, negli ultimi tre lustri la disponibilità di grandi corpora elettronici ha completamente mutato il panorama degli studi sul lessico, oltre che i metodi e il modo stesso di pensare la ricerca in linguistica e probabilmente in tutte le scienze cosiddette empiriche. Ciò vale, anche se finora solo in misura limitata, anche per la prospettiva diacronica, per la quale cominciano solo di recente a essere disponibili corpora elettronici (cfr. una prima rassegna in CLARIDGE 2008). In quest'orizzonte si iscrive anche il corpus MIDIA (= *Morfologia dell'Italiano in DIACRONIA*), il quale, pur essendo di grandezza limitata a circa 7,5 milioni di *tokens*, è – sia detto con un po' d'enfasi – il più grande corpus bilanciato per tipi di testo ed epoche storiche rappresentate disponibile per una lingua romanza, e non solo. Come si evince dalla Tab. 1¹, per lo spagnolo il RAE e il MD-S sono di taglia molto più grande, rispettivamente di ca. 250 e 100 milioni di *tokens*, ma non sono bilanciati, mentre l'etichettatura automatica per parti del discorso (il cosiddetto *POS-tagging*) è solo parziale. Un discorso analogo vale per il MD-P del portoghese e il CICA del

* Università di Torino.

** Parti di questo contributo sono state presentate al 21st *International Conference on Historical Linguistics* svoltosi a Oslo (6.-9.8.2013), al 16th *International Morphology Meeting* svoltosi a Budapest (29.5.-1.6.2014), all'11th *International Conference on Actionality, Tense, Aspect, Modality/Evidentiality (Chronos 11)* svoltosi presso la Scuola Normale Superiore di Pisa (16.-18.6.2014). Ringrazio Marco Angster e Claudio Iacobini per aver elaborato e discusso con me alcuni aspetti di queste ricerche, oltre che i colleghi e gli amici presenti ai convegni per osservazioni e commenti. Ovviamente, la responsabilità di quanto è contenuto in queste pagine, inclusi errori e sviste, è solo mia.

¹ Tra le lingue romanze manca ogni riferimento al francese, per il quale l'unico corpus diacronico di una certa rilevanza, FRANTEXT, che dichiara di essere «la seule à proposer des recherches sur des textes qui vont de 1180 à 2012», non è accessibile liberamente e pertanto non viene discusso in questa rassegna. In ogni caso, pur se la taglia è apparentemente molto ampia al punto da coprire l'intera storia linguistica del francese, il corpus FRANTEXT non sembra essere bilanciato.

catalano. Infine, dalla Tab. 1 si evince che anche il confronto con l'inglese, lingua in genere all'avanguardia per studi e materiali disponibili come testimonia CLARIDGE 2008, è incoraggiante. Infatti, mentre esistono corpora diacronici di taglia più ampia ma non bilanciati come l'OED-MD, l'HC, che è l'unico corpus diacronico bilanciato disponibile, è decisamente più piccolo di MIDIA.

	RAE	MD-S	MD-P	CICA	OED-MD	HC	MIDIA
Grandezza (tokens)	237 M	100 M	45 M	9 M	57 M	1,5 M	7 M
Bilanciamento	no	no	no	no	no	sì	sì
POS-tagging	no	sì	sì	sì/no	sì (basico)	sì	sì
Estensione temporale (sec.)	XI-XX	XIII-XX	XIV-XX	XI-XVIII	X-XX	VIII-XVIII	XIII-XX

Tab. 1: Corpora diacronici accessibili delle lingue romanze e dell'inglese.

Insomma, benché più piccolo di altri corpora diacronici, MIDIA è integralmente etichettato e copre tutto l'arco temporale della storia linguistica dell'italiano. Inoltre a parte il fatto che è più grande dell'HC, MIDIA è bilanciato sia per epoche storiche che per generi testuali, mentre l'HC è bilanciato solo rispetto alle epoche storiche coperte.

Se spostiamo il confronto ai corpora disponibili per l'italiano (su cui si veda BARONI 2010 per una rassegna), i risultati sono di nuovo incoraggianti, come si vede dalla Tab. 2:

	OVI	<i>La Stampa</i>	<i>la Repubblica</i>	ItWaC	CoLFIS	MIDIA
Grandezza (tokens)	23 M	75 M	325 M	1,5 B	3,8 M	7 M
Bilanciamento	no	no	no	no	sì	sì
Estensione temporale	-1375	1996-1998	1985-2000	1992-1994	XX sec.	-1945

Tab. 2: Confronto tra corpora disponibili dell'italiano.

Benché sia molto più piccolo di altri corpora sia diacronici come l'OVI, sia sincronici come il *Corpus «La Stampa»* (cfr. GAETA/RICCA 2002; 2003) e il *Corpus «la Repubblica»* (cfr. BARONI *et al.* 2004) o un corpus che raccoglie testi dal web come l'ItWac (cfr. BARONI *et al.* 2009), tuttavia rispetto a questi ultimi è pienamente bilanciato. Infine è di grandezza doppia rispetto all'unico corpus bilanciato dell'italiano, il CoLFIS (cfr. LAUDANNA *et al.* 1995), che copre tuttavia solo l'italiano attuale. In buona sostanza, MIDIA cerca di combinare il meglio di entrambe le prospettive: è perfettamente bilanciato (per epoche e tipi di testo) ed è abbastanza grande (anzi è *più* grande dei corpora diacronici e sincronici bilanciati disponibili).

Dopo aver discusso brevemente la prospettiva aperta dai corpora testuali nelle ricerche sul lessico (§ 2), in quanto segue mi propongo di indagare l'affidabilità del corpus MIDIA per mezzo di un paio di studi connessi con i nomi deverbali che mi sono noti grazie a precedenti ricerche, in particolare i nomi d'agente maschili e femminili (§ 3) e i cosiddetti derivati con la forma del participio passato femminile (§ 4). L'interesse primario dell'indagine consiste, oltre che nel mostrare l'utilità di MIDIA per ricerche di questo tipo, nel verificare la consistenza in prospettiva evidentemente diacronica di questi due procedimenti derivazionali, e di riflettere in questa maniera sul valore delle ricerche basate su corpora per l'indagine della formazione delle parole dal punto di vista diacronico. Il § 5 tira brevemente le conclusioni dell'indagine.

2. Dinamica lessicale e formazione delle parole

L'approccio tradizionale allo studio della dinamica lessicale, e in particolare a quella connessa con la formazione delle parole, è fondamentalmente basato sull'uso dei dizionari. I pregi e i difetti di quest'approccio sono noti e sono stati ampiamente discussi (cfr. per l'italiano GAETA/RICCA 2002; 2003; 2005 e più in generale BAAYEN 1989; PLAG 1999; BAUER 2001). Per soffermarci solo sugli aspetti più strettamente connessi con la produttività, i dati presi in considerazione nei dizionari fanno ovviamente riferimento ai lemmi, cioè ai *types*, estratti da basi testuali di solito non specificate, che possono essere considerate "virtuali" nella misura in cui il compilatore del dizionario fa riferimento implicito – nel migliore dei casi – a una serie di testi dai quali estrae i lemmi e che poi non sempre vengono esplicitati in bibliografia. Anche in questi casi fortunati, tuttavia, i diversi tipi di testo non vengono considerati se non marginalmente e viene trascurato completamente ogni riferimento ai *tokens*. In quest'ottica, le nuove formazioni, che segnalano la produttività di un certo procedimento derivazionale, sono dei nuovi *types* non attestati in precedenza nei dizionari. In seguito alla rivoluzione dei corpora è diventato rilevante, oltre al numero dei *types*, anche quello dei *tokens* estratti da basi testuali concrete. Ciò ha permesso di valutare la significatività del genere testuale e della grandezza del corpus (PLAG *et al.* 1999) per cui le nuove formazioni sono dei *types* non attestati in precedenza in determinate porzioni testuali sufficientemente ampie: i cosiddetti *hapax legomena* (BAAYEN 1989; 2009).

Tra le varie obiezioni che sono state sollevate contro l'approccio basato su corpora (cfr. BAUER 2001; GAETA/RICCA 2006; 2015) una è particolarmente rilevante per la prospettiva diacronica adottata qui: gli *hapax legomena* sono una misura affidabile della produttività di un certo procedimento di formazione delle parole solo in corpora sufficientemente ampi. Infatti, è solo dopo che si è raccolto un rilevante numero di *tokens* che possiamo affermare con una certa sicurezza che non ci troviamo di fronte semplicemente a una parola di fascia di frequenza medio-bassa.

Lavorando su un corpus sincronico di ca. 75 milioni di *tokens*, GAETA/RICCA 2006 si affidano a porzioni testuali non inferiori ai 6 milioni di *tokens*, proprio perché la frequenza 1:6000000 è ben al di sotto della frequenza anche bassa in genere associata con una parola presente nel nostro lessico mentale (cfr. BAAYEN 1994 per una stima). È chiaro che queste procedure non sono immediatamente replicabili nel nostro caso, dato che la taglia complessiva di MIDIA è di 7,5 milioni di *tokens*.

Per questi e altri motivi, BAUER 2001 suggerisce una soluzione di compromesso che si basa sul confronto sistematico di valori estratti sia dai dizionari sia dai corpora, che dovrebbe essere particolarmente adatta per la prospettiva diacronica. Per quanto la soluzione suggerita da Bauer sembri interessante e certamente degna di esser verificata sperimentalmente, non ha avuto grosso seguito. Probabilmente, a giocare contro quest'idea è la sua macchinosità rispetto invece alla pratica proposta di BAAYEN 2009 di impiegare come indice di produttività meramente il numero di *hapax legomena* presenti in un corpus di estensione sufficientemente ampia. In effetti, nonostante le obiezioni di carattere concettuale che sono state sollevate contro la cosiddetta *hapax-conditioned productivity* (o P^*), in realtà nella maggioranza dei casi essa correla con l'indice di produttività h/N (o P), che misura la probabilità di incontrare un nuovo *type* h formato con un certo affisso dopo aver computato un numero N di *tokens* di quell'affisso. Nonostante il fatto che il numero di h di per sé non misuri questa probabilità P , tuttavia P^* risulta empiricamente connessa con essa quasi sempre, per cui viene utilizzato come indicatore – per quanto impreciso – di produttività².

Inoltre, per la prospettiva diacronica adottata qui è utile far riferimento a un'altra caratteristica che si osserva negli h : si può infatti mostrare empiricamente che è tra gli h che si annidano le parole rare, che per definizione sono quelle che non vengono riportate dai dizionari (cfr. GAETA/RICCA 2002 per un'indagine a proposito di due dizionari di taglia rispettivamente media e massimale come *Zingarelli* e GRADIT). Questa caratteristica è molto utile in prospettiva diacronica, proprio perché tradizionalmente la produttività in diacronia viene misurata a partire da quanto viene riportato dai dizionari. Si aggiunga che MIDIA è un corpus, seppur di taglia limitata, tuttavia bilanciato per epoche indagate e tipi di testo. Pertanto, ambisce a fornire dati che in qualche misura permettano di alleviare almeno in parte la distorsione che risulta dall'uso incontrollato di corpora, per di più di taglia e dimensioni difficilmente comparabili per epoche e generi testuali come si è visto in Tab. 1 – cosa che avviene spesso per la diacronia di una lingua (cfr. COWIE 2000;

² In particolare, in GAETA 2007 è stato mostrato che i valori di P e P^* divergono in maniera sistematica quando si confrontano le produttività di affissi che presentano un rango di frequenza piuttosto differente. In questi casi, il valore di P^* è sistematicamente più alto nel caso degli affissi più frequenti, nonostante che i valori di P – ottenuti secondo la procedura a corpus variabile che tiene appunto sotto controllo la frequenza degli affissi nel corpus complessivo – siano sostanzialmente analoghi.

COWIE/DALTON-PUFFER 2002 per una riflessione sul concetto di produttività e sui fattori in gioco in prospettiva diacronica). Per questi motivi, i dati necessariamente limitati che MIDIA offre possono rivelare spunti di analisi interessanti proprio per le qualità del corpus. In quanto segue cercherò di verificarne l'affidabilità, indagando brevemente alcuni procedimenti di formazione delle parole e comparandone i risultati con quanto è possibile ottenere dai corpora sincronici.

3. Qualche indagine con MIDIA: frequenze, h, distribuzioni, allomorfie

Per saggiare l'affidabilità di MIDIA utilizzerò innanzitutto la coppia di suffissi d'agente maschile e femminile che ritroviamo in parole come *nuotatore* e *nuotatrice* (cfr. GAETA 2010; 2012). Per cominciare, si può confrontare la frequenza assoluta dei derivati con questi suffissi e il *Corpus «La Stampa»* (cfr. GAETA/RICCA 2002; 2006) che contiene 75 milioni di *tokens*:

	<i>La Stampa</i> (1996-98) – 75 M				MIDIA			
	N	V	h	F	N	V	h	F
-(t)ore	273706	1480	461	3.7%	22649	836	270	3.6%
-trice	23780	645	224	0.3%	1341	311	150	0.2%

Tab. 3: I nomi d'agente nel *Corpus «La Stampa»* e in MIDIA.

Pur se i valori assoluti del numero di *tokens* (= N) e *types* (= V), oltre che di *bapax legomena* (= h) divergono notevolmente, le frequenze assolute (= F) nei due corpora collimano e riflettono anche la nota grande differenza tra il suffisso maschile e quello femminile. Passiamo ora al valore di h, che evidentemente diverge data la taglia complessiva dei due corpora. Tuttavia, nonostante la piccola taglia del corpus, si riproduce lo stesso effetto osservato con il *Corpus «La Stampa»* (cfr. GAETA/RICCA 2002), come si vede nella Fig. 1:

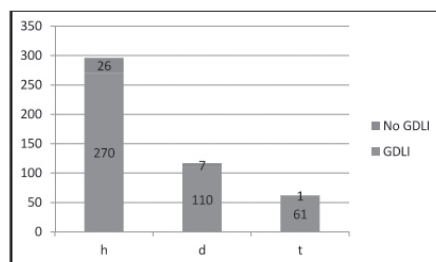


Fig. 1: Valori di *bapax*, *dis* e *tris legomena* formati con il suffisso *-tore* estratti da MIDIA e presenti nel GDLI.

Se teniamo presente i derivati con frequenza 1 (= *h*), 2 (= *d*) e 3 (= *t*) formati con il suffisso *-tore*, è tra gli *h* che si annidano le parole che non sono attestate nel *thesaurus* dell'italiano, il GDLI. Queste ammontano a circa il 10% degli *h* e sono riportate qui di sotto senza ulteriori commenti:

accimatore	coaiutatore	gherbellatore	pannatore
ammorbatore	coitatore	giustiziatore	pignoratore
assetatore	compromettitore	imborsatore	progettatore
astensore	confederatore	imbrazatore	restatore
avantatore	confettore	imbussolatore	restrittore
balitore	contrattore	impresore	scancellatore
biancatore	disaccatore	infintore	scrocchiatore
cervolatore	dissimulatore	inzuppatore	soscrittore
coadiutatore	fidatore	lassatore	staggitore

Evidentemente, queste parole sono pronte per arricchire il patrimonio lessicale di dizionari "massimali" come il GDLI o il GRADIT³.

Verificata la generale affidabilità di MIDIA rispetto alla frequenza assoluta dei derivati e al valore degli *h*, possiamo ora studiare l'andamento diacronico di questi ultimi, cioè la distribuzione longitudinale. Nelle Figg. 2 e 3 si osserva la distribuzione dei valori in termini di *V* e di *h* presenti nel corpus, sia con il suffisso *-tore*

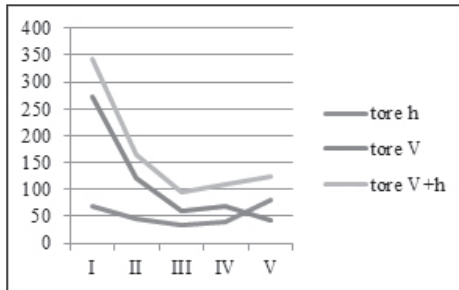


Fig. 2: Distribuzione longitudinale del suffisso *-tore*.

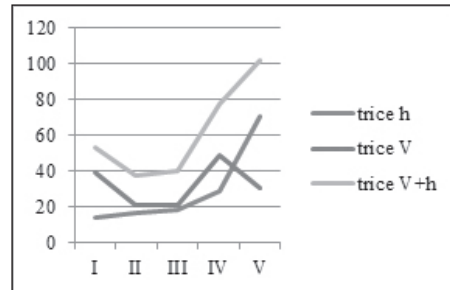


Fig. 3: Distribuzione longitudinale del suffisso *-trice*.

³ Il confronto con l'analogo femminile *-trice* è impossibile perché nel GDLI non è per lo più individuato come entrata lessicale autonoma, ma è segnalato all'interno del corrispettivo lemma maschile in *-tore* per mezzo della formula: «femm. *-trice*», che per altro solleva nel lettore il dubbio se il femminile sia effettivamente attestato o se sia meccanicamente indicato dal compilatore.

sia con il femminile *-trice*. Nei grafici, confrontando il dato di V con e senza h, si osserva la produttività dei due suffissi:

La differenza di numerosità tra i due suffissi rende diversi i due grafici perché il valore V di *-tore* scorporato dagli h è costantemente più alto di quest'ultimo nei primi quattro periodi, cioè dalle origini al 1375 (= I), dal 1376 al 1532 (= II), dal 1533 al 1691 (= III) e dal 1692 al 1840 (= IV). Tuttavia, nell'ultimo periodo, dal 1841 al 1947 (= V), i due valori si incrociano per entrambi i suffissi, testimoniando la grande produttività che si registra negli ultimi secoli. Nelle Figg. 4 e 5 si con-

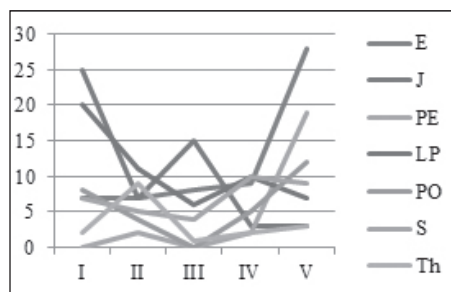


Fig. 4: Distribuzione degli h formati con il suffisso *-tore* per tipi di testo.

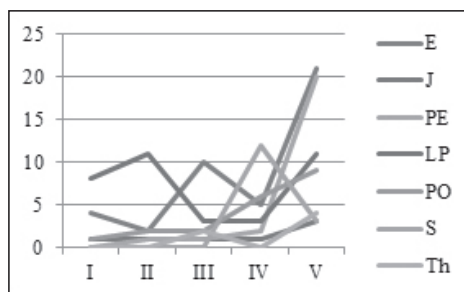


Fig. 5: Distribuzione degli h formati con il suffisso *-trice* per tipi di testo.

frontano invece i dati riguardanti gli h dei due suffissi distribuiti in base al genere testuale:

Benché – di nuovo – la caratterizzazione sia del tutto superficiale, il confronto è interessante perché mostra come i due suffissi presentino un andamento parzialmente analogo con una crescita consistente negli ultimi secoli per quanto riguarda il genere dei testi espositivi (E) e personali (PE), e in parte in quelli poetici (PO), una netta differenza a favore del suffisso maschile – almeno nei primi secoli – nei testi normativi (J) e invece a favore di quello femminile – soprattutto negli ultimi secoli – nella prosa letteraria (LP), mentre molto più oscillante è il dato riguardante gli h nei testi scientifici (S) e in quelli più vicini al parlato come i testi teatrali (Th). Si tratta evidentemente di microtendenze, che sono rilevanti nella misura in cui il corpus è bilanciato, e che in parte corrispondono a fenomeni noti come l'esplosione dei nomi d'agente nel Sette-Ottocento e in parte necessitano di ulteriori indagini mirate sui vari generi testuali.

Infine, dopo aver discusso la questione di cos'è un nuovo *type* formato con un certo suffisso mettendo in evidenza il ruolo degli h, diamo velocemente uno sguardo alle questioni connesse con l'allomorfia della base, chiedendoci – in maniera complementare alla domanda di cos'è una nuova formazione – cos'è produttivo come base di un procedimento di formazione delle parole.

Com'è noto, per i procedimenti deverbali, abbiamo diverse possibilità che sono state interpretate in maniera differente (cfr. GAETA 2002: 63-69 per la ras-

segna). Le quattro possibilità sono: la base morfologica del tema verbale (= TV), un “morfoma” nel senso di ARONOFF 1994, in genere formalmente identico con l'imperativo con l'esclusione del suffisso *-isc-*, quella del participio passato (= PP), quella del participio passato latino (= PPL), che è una base morfologica addizionale giustificata solo in termini paradigmatici, e quella della radice verbale (= Rad) priva della vocale tematica:

- (1)
- | | | |
|----------------------|--|---|
| a. <i>rapitore</i> | [[<i>rapi</i>] _{TV} - <i>tore</i>] | [[<i>rapit</i>] _{PP} - <i>ore</i>] |
| b. <i>scopritore</i> | [[<i>scopri</i>] _{TV} - <i>tore</i>] | *[[<i>scopert</i>] _{PP} - <i>ore</i>] |
| c. <i>direttore</i> | *[[<i>dirigi</i>] _{TV} - <i>tore</i>] | [[<i>dirett</i>] _{PP} - <i>ore</i>] |
| d. <i>seduttore</i> | *[[<i>seduci</i>] _{TV} - <i>tore</i>] | *[[<i>sedott</i>] _{PP} - <i>ore</i>] [[<i>sedutt</i>] _{PPL} - <i>ore</i>] |
| e. <i>recensore</i> | *[[<i>recensi</i>] _{TV} - <i>tore</i>] | *[[<i>recensit</i>] _{PP} - <i>ore</i>] [[<i>recens</i>] _{Rad} - <i>ore</i>] |

Mentre i casi in (1d-e) sono univoci e fanno riferimento in (1d) a un meccanismo derivazionale opaco sincronicamente perché formato su una base che fa riferimento al morfoma di participio passato latino e in (1e) al morfoma costituito dalla radice lessicale chiaramente improduttivo, il caso (1a) è ambiguo tra l'interpretazione di (1b) in cui il nome d'agente è formato sul morfoma del tema verbale – preferita da THORNTON 1990, 1991 – e l'interpretazione in (1c) in cui la base è data dal morfoma di participio passato, come propone SCALISE 1983. Come base di confronto sincronica si userà di nuovo il *Corpus «La Stampa»* (GAETA/RICCA 2002; 2006) che consiste di 75 milioni di *tokens*:

	<i>La Stampa</i> 1996-98			MIDIA		
	N	V	h	N	V	h
TV/PP	145356 52%	1256 85%	421 91%	10531 47%	600 72%	213 80%
solo TV	15978 6%	57 4%	17 4%	2083 9%	98 12%	32 12%
solo PP	50219 18%	58 4%	7 2%	3161 14%	42 5%	13 5%
PPL	50115 18%	89 6%	13 3%	5604 25%	75 9%	6 2%
Rad	19495 7%	21 1%	3 1%	1185 5%	13 2%	2 1%
Tot.	281073	1481	461	22564	828	266

Tab. 4: L'allomorfia dei nomi d'agente nel *Corpus «La Stampa»* e in MIDIA.

Il dato interessante è la maggiore preferenza in MIDIA per la selezione univoca del tema verbale, circa tre volte superiore rispetto al *Corpus «La Stampa»* sia per gli h che i per *types* V. Si noti inoltre come rispetto al *Corpus «La Stampa»* sia decisamente più basso in MIDIA – circa il 10% in meno sia per gli h che per i V –

il numero di derivati ambigui tra l'analisi in termini di tema verbale e di participio passato. Insomma, dalla prospettiva diacronica offerta da MIDIA sembra emergere una consistente preferenza per il morfoma di tema verbale rispetto a quello di participio passato che si riscontra, seppur in misura leggermente inferiore, anche in sincronia (cfr. GAETA 2002: 72-73).

4. I nomi deverbali in forma di participio passato femminile

L'altro caso che intendo discutere con l'ausilio di MIDIA riguarda i derivati deverbali che coincidono formalmente con il participio passato femminile: il tipo *mangiata* e *corsa* (cfr. GAETA 2000; 2002), e quelli denominali formati con il suffisso *-ata* che storicamente derivano dai primi: il tipo *coltellata* e *pagliacciata*. Il contributo empirico che MIDIA può offrire per questi derivati è notevole, in quanto permette di individuare con relativa facilità i nomi deverbali rispetto ai meri participi grazie alla ricerca per categoria lessicale, eventualmente in un contesto sintatticamente specificato. Data la novità del prodotto, è la prima volta che è possibile estrarre questo tipo di dati almeno rispetto alla dimensione diacronica⁴. Prima di presentare i dati, tuttavia, è necessario richiamare brevemente le caratteristiche peculiari dei nomi deverbali formati con il participio passato femminile (= FPP). Non mi soffermerò affatto sulle caratteristiche formali, cioè se si debba partire da un morfoma di participio passato cui affiggere un morfema *-a*, o se si tratti di una conversione che seleziona il femminile a partire dal morfoma di participio passato, o infine se si tratti di un suffisso *-ata* che seleziona sia verbi che nomi cui si affianca un ampio numero di allomorfie simili a quelle viste sopra a proposito dei nomi d'agente maschili (cfr. GAETA 2002: 141-145; 2004 per una discussione). Invece, mi concentrerò esclusivamente sulle proprietà semantiche dei derivati che chiamerò genericamente (deverbali) FPP.

Da questo punto di vista, osserviamo innanzitutto che i nomi deverbali in italiano sono ben distribuiti rispetto alla selettività azionale dei rispettivi procedimenti di formazione, che possono essere di natura morfologica con riferimento al tipo FPP (2d), ma anche al tipo formato con i suffissi *-mento* (o *-zione*) (2a), o con il cosiddetto suffisso *-(z)a* che si combina con il morfoma di participio presente (2b), o infine al tipo formato per nominalizzazione sintattica, cioè il cosiddetto infinito sostantivato INF (2c) (cfr. GAETA 2002: 109-121; 2004 per la rassegna):

⁴ Non è possibile questa volta confrontare i dati di MIDIA con quelli del *Corpus «La Stampa»*, perché quest'ultimo non è etichettato e dunque non permette di separare i participi dai nomi. Inoltre, nonostante la disponibilità di altri corpora sincronici etichettati come il *Corpus «la Repubblica»*, non sono state fatte – almeno per quanto ne so io – indagini mirate su questo procedimento di formazione delle parole. Pertanto i dati presentati qui sono una novità assoluta, ancorché inquadrati in una cornice diacronica.

- (2) a. $[[\dots]_{Vi} -mento]_{Nj} \leftrightarrow [\text{the SitTyp involved by SEM}_i \dots]_j$
*conseguire*₁ → *conseguimento*
Il conseguimento del titolo di direttore fu ottenuto in due anni.
- b. $[[\dots <-dyn>]_{Vi} -(z)a]_{Nj} \leftrightarrow [\text{the SitTyp involved by SEM}_i \dots]_j$
*conseguire*₂ → *conseguenza*
**La conseguenza della sconfitta dal suo errore fu ottenuta in due anni.*
- c. $[[\dots]_{Vi} INF]_{Nj} \leftrightarrow [\text{the SitTyp } <-b>_j \text{ combined with SEM}_i \dots]_j$
*conseguire*₁ → *conseguire*
**Il conseguire il titolo di direttore fu ottenuto in due anni.*
- d. $[[\dots <+dyn, -b>]_{Vi} FPP]_{Nj} \leftrightarrow [\text{the SitTyp } <+b>_j \text{ combined with SEM}_i \dots]_j$
dormire, nuotare → *dormita, nuotata*
costruire, formare → **costruita, *formata*

Per la rappresentazione mi sono servito – con le opportune modifiche che spiegherò qui di sotto – della notazione proposta da BOOIJ 2010 nel suo modello di morfologia costruzionale, anche se in realtà da un lato la notazione non è particolarmente vincolante dal punto di vista teorico ed è stata scelta apposta per questo, e dall’altro la rappresentazione è completamente indipendente dalla sua notazione⁵. In ogni caso, la notazione prevede due parti vincolate in modo biunivoco. Nella parte sinistra è indicata la struttura formale dell’unità lessicale cui è associata la parte semantica alla sua destra. Si noti che la notazione abbrevia una struttura più complessa perché contiene l’input, cioè un verbo qualsiasi $[\dots]_{Vi}$ non ulteriormente specificato dal punto di vista delle proprietà semantiche e coincidentizzato con una certa semantica $SEM_i \dots$ che si ritrova nella parte di destra. (2a) va allora interpretato così: il suffisso *-mento* seleziona un qualsiasi verbo e ne costruisce il nome d’azione $[\dots]_{Nj}$, che descrive cioè il *situation type* (= SitTyp) che rispecchia le proprietà semantiche della base $SEM_i \dots$ con l’aggiunta della semantica portata dal suffisso e coincidentizzata con esso.

Ma qual è questa semantica? Nel caso di questo suffisso è molto “leggera” al punto da non presentare effetti particolari sul SitTyp di input, a parte quanto discende dal fatto stesso di formare un nome, cioè il tipico effetto di “reificazione” del predicato (cfr. TALMY 1988: 175; GAETA 2002: 29)⁶. Gli altri casi esemplificano la diversa tipologia di selezione dell’input e di modificazione dell’output. In (2b) il suffisso *-(z)a* seleziona tipicamente predicati stativi (cioè non dinamici: *<dyn>*) che mantengono la loro proprietà di atelicità, come mostra l’esempio riportato, in cui *conseguenza* derivato dallo stativo *conseguire*₂ non è compatibile con un predicato contenitore che forza l’elicitazione del punto finale del SitTyp descritto dal verbo, come si vede dal contrasto con *conseguimento* in (2a) che deriva dal trasformativo

⁵ Per una rappresentazione per mezzo della Grammatica Cognitiva si veda GAETA 2002: 178-185; 2003.

⁶ L’effetto di reificazione si osserva nella presenza di alcune proprietà tipiche dei nomi come la pluralizzabilità che nel caso dei nomi d’azione prende tra l’altro la veste della plurazionalità (cfr. GAETA 2015).

*conseguire*₁⁷. Nel tipo (2c) vengono invece modificate le proprietà di output della nominalizzazione, che perde il tratto di telicità (o *boundedness*: <-b>) nel caso esso sia presente nel verbo di input come in *conseguire*₁, che diviene incompatibile con il predicato contenitore che forza la telicità. Infine, (2d) è il caso del FPP, che è selettivo rispetto alle proprietà del predicato sia in input, che deve essere dinamico e atelico per cui non sono possibili formazioni a partire da verbi risultativi come **costruita* e **formata*, sia di output, perché il nome risultante è invece telico come si mostra dalla compatibilità con predicati contenitore che forzano la descrizione dello stato risultante: *La nuotata di Gianni si è conclusa dopo due ore con una doccia*. Più in particolare, il FPP ha l'effetto di un *packaging operator* (operatore di "impacchettamento") che è noto nella letteratura semantica a proposito di quelle espressioni che servono a quantificare nomi di massa, che in genere sono rappresentati come una sostanza non delimitata e non granulare (cfr. JACKENDOFF 1991; GAETA 2000; 2002: 145-148; 2003). Il *packaging operator* serve proprio a isolare singole parti di una massa indistinta come *un pugno di sabbia* rispetto a *sabbia* o *un(a tazza di) caffè* rispetto a *caffè*. In termini grafici si ottiene la figura seguente⁸:

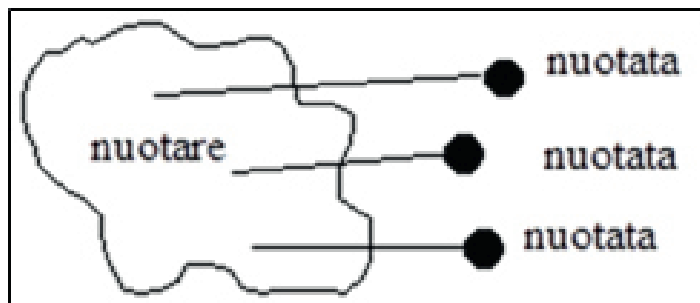


Fig. 6: Il FPP come operatore di *packaging*.

Le singole istanziazioni di *nuotare* sono individuate (e rese pluralizzabili) dal FPP *nuotata*. In questa funzione di *packaging operator*, il FPP è usato in particolare nella costruzione con i verbi supporto *fare* e *dare*, rispettivamente con basi intransitive (3a) e transitive (3b) (cfr. GAETA 2002: 150-163; 2004):

⁷ Per inciso, adotto la terminologia di BERTINETTO 1986, che, com'è noto, a sua volta adatta le quattro classi azionali di VENDLER 1967 in stativi (*stative*), continuativi (*activity*), risultativi (*resultative*) e trasformativi (*achievement*). Per quanto riguarda i predicati contenitore si veda BARTSCH 1986 e la discussione in GAETA 2002: 116-117.

⁸ È possibile anche il procedimento opposto che richiede un *grinding operator* (operatore di "frantumazione"), come nell'espressione: *In quest'insalata c'è dell'ananas*, in cui l'entità *ananas* che di per sé è delimitata e pluralizzabile viene "frantumata" in massa non granulare.

- (3) a. *Guido ha fatto una bella dormita / nuotata / mangiata / camminata.*
 b. *Gianni ha dato una piegata / stirata / lavata / pulita alle camicie.*
 c. *Gianni ha dato una piegata veloce / *accurata alle camicie.*

Si noti che nella perifrasi con verbo supporto il SitTyp denotato dal FPP è rappresentato come svolto in modo sbrigativo e impreciso ancorché chiaramente concluso (3c).

Infine, il FPP può essere considerato un “rivelatore” di trasformatività perché è ampiamente compatibile con i verbi trasformativi, nei quali però è in genere l’unico nome d’azione possibile mentre non trova sorprendentemente applicazione il tipo in (2a) che come si ricorderà presenta una bassa selettività e quindi ampia compatibilità in input:

- (4) *cadere* → *caduta* / **cadimento* / **cadizione*
cacciare → *cacciata* / **cacciamento* / **cacciazione*
entrare → *entrata* / **entramento* / **entrazione*
fermare → *fermata* / **fermamento* / **fermazione*
salire → *salita* / **salimento* / **salizione*
uscire → *uscita* / **uscimento* / **uscizione*
venire → *venuta* / **venimento* / **venizione*

Ciò si spiega grazie al valore complessivo del FPP illustrato dalla sua descrizione semantica in (2d) riportata qui di sotto:

- (5) $[[\dots <+dyn, -b>]_{vi} FPP]_{Nj} \leftrightarrow [\text{the SitTyp } <+b>_i \text{ combined with SEM}_i \dots]_j$

La parte in grassetto in (5) identifica proprio i verbi trasformativi che sono caratterizzati dalla telicità oltre che dalla dinamicità, per cui lo schema in (5) ne guida la derivazione in base alla perfetta corrispondenza dell’output, cioè per mezzo di una *output-driven derivation*. Quest’ultima è per altro connaturata all’approccio costruzionale di BOOIJ 2010 nella misura in cui lo schema costruzionale biunivoco stabilisce rapporti paradigmatici piuttosto che associare un input con un output come nell’approccio basato su regole à la ARONOFF 1976.

Rispetto a questo quadro complesso, l’origine diacronica del FPP è piuttosto oscura. In letteratura sono rappresentate due ipotesi genetiche diverse (cfr. GAETA 2002: 142; FABRIZIO 2015 e FABRIZIO in stampa per una rassegna): l’una ricostruisce un processo di nominalizzazione a partire dal femminile singolare del participio passato dopo l’ellissi del nome testa: (PECUNIA) COLLECTA > *colletta*, (RES) EXPENSA > *spesa*, ecc. L’altra invece parte dal neutro plurale dei nomi d’azione latini formati sulla base del participio passato, in seguito rianalizzati come femminili singolari: ACCESSU(M) / ACCESSA, OFFENSU(M) / OFFENSA, ecc. Apparentemente la prima ipotesi è illuminante perché, oltre a far ricorso a un fenomeno piuttosto diffuso nella creazione di *pattern*

lessicali ripetuti come l'ellissi⁹, gli esempi sembrano particolarmente ben spiegati. In realtà, quest'ipotesi genetica non è facilmente generalizzabile a meno di individuare ben definiti *pattern* lessicali, sostanzialmente indipendenti dalla specifica semantica verbale, ma connessi con alcuni nomi testa particolarmente salienti. Sicché, a parte i casi indicati, non sembra molto chiaro come possa essersi generalizzato il modello morfologico, che invece abbiamo visto essere piuttosto ben individuato dal punto di vista della selettività in entrata e della *facies* semantica in uscita.

Nella seconda ipotesi, invece, la semantica verbale ha un ruolo primario perché – come vedremo fra un momento – i nomi d'azione latini in questione presentavano precise capacità selettive, e in particolare è possibile individuare dei chiari paralleli tra il modello latino in *-tus* riguardo alla selettività dei predicati e il FPP (cfr. FABRIZIO 2015). La *facies* semantica dei deverbali FPP sembra pertanto parlare a favore di questa seconda ipotesi.

Ma lasciando da parte per il momento le questioni di natura genetica, riguardo alla diacronia dell'italiano si possono formulare almeno due ipotesi diverse che fanno predizioni testabili empiricamente. Da un lato, data la sua specializzazione per la telicità (o *boundedness*) di uscita del nome deverbale, ci si può chiedere se la funzione di operatore di *packaging* sia antica. Se ciò fosse vero, ci si dovrebbe attendere un'estensione diacronica ai trasformativi a partire da un input iniziale costituito per lo più da continuativi. L'estensione sarebbe motivata dalla corrispondenza perfetta dei trasformativi con la *facies* semantica dei FPP.

D'altro canto, la funzione di operatore di *packaging* potrebbe essersi sviluppata a partire dal gruppo di nomi d'azione formati da trasformativi, e di lì essersi estesa ai continuativi introducendo il valore, semelfattivo, di *packaging*. In questo secondo caso, ci dovremmo attendere invece il quadro opposto: già in antico il FPP fungerebbe da rivelatore di trasformatività, mentre la funzione di operatore di *packaging* si sarebbe sviluppata successivamente generalizzando la *facies* semantica tipica dei trasformativi ai predicati continuativi.

Per dirimere la questione, diamo un'occhiata ai dati estraibili da MIDIA. Per mezzo delle ricerche consentite dai campi riguardanti la classe di parola è possibile (non senza un certo lavoro inevitabile di ripulitura manuale dei dati) arrivare al grafico della Fig. 7:

⁹ RAINER 2004a, per esempio, sottolinea come l'ellissi del nome testa abbia avuto un ruolo essenziale nella diffusione dei derivati per mezzo del suffisso dei nomi d'agente, specialmente in ben definite serie lessicali come (*muscolo*) *adduttore* / *rotatore*, (*numero*) *fattore* / *moltiplicatore*, ecc.

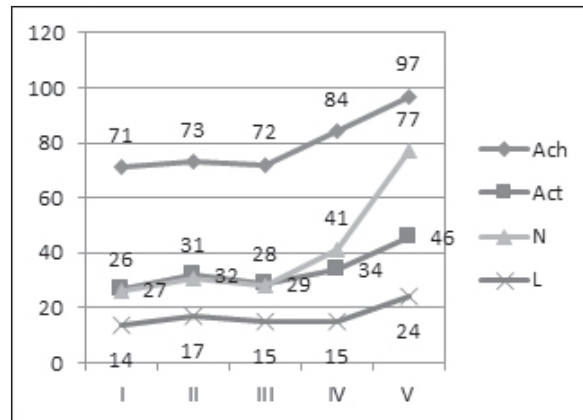


Fig. 7: I derivati da FPP e denominali in MIDIA.

I dati mostrano che i derivati da FPP formati a partire da verbi trasformativi (= *Ach*) rappresentano il nucleo di gran lunga più numeroso, seguito a una notevole distanza dai derivati da verbi continuativi (= *Act*). Questi ultimi sono appaiati ai derivati formati a partire da basi nominali (= *N*) – sui quali tornerò tra un attimo – e seguiti dal gruppo di formazioni eterogenee che sono intese genericamente come lessicalizzate (= *L*), cioè opache. Non mi soffermerò su queste ultime: basti dire che sono casi in cui non è rintracciabile la semantica esemplificata sopra (*detta*, *tinta*) e in alcuni casi anche il rapporto con il verbo base non è perspicuo o possibile (*data*, *colletta*)¹⁰.

La seconda ipotesi sembra pertanto uscire corroborata da questi numeri rispetto alla prima: il tipo FPP è robustamente ancorato alla sua funzione primaria di rivelatore di trasformatività. La funzione di operatore di *packaging* messa in atto con predicatori continuativi è secondaria, ma cresce in maniera molto significativa, soprattutto in epoca recente. Come si evince dalla Tab. 5, l'incremento registrato con basi continuative dal primo al quinto periodo è di gran lunga più significativo rispetto a quello registrato con basi trasformative:

Base	$\Delta I > V$
Verbi trasformativi (= Ach)	+ 37%
Verbi continuativi (= Act)	+ 70%
Nomi (= N)	+ 196%

Tab. 5: L'incremento dei derivati FPP rispetto al tipo di base.

¹⁰ Per una discussione di questi tipi nell'italiano odierno si veda GAETA 2002: 168-171.

Questo secondo me è il dato più significativo perché sgombera il terreno da una possibile obiezione legata ai limiti di taglia di MIDIA. Si potrebbe infatti obiettare che non sono i derivati da trasformativi a essere più numerosi, ma semplicemente che questi ultimi sono più frequenti e quindi intercettati da MIDIA. L'obiezione cade però tenendo presente la grande espansione delle basi continuative soprattutto in epoca recente: a meno di immaginare grosse oscillazioni di frequenza, si osserva evidentemente nell'Otto-Novecento un indubbio incremento di derivati sia da basi trasformative che da basi continuative, con una chiara prevalenza di queste ultime. Insomma, mi sembra che si possa concludere che la funzione di operatore di *packaging* si sia sviluppata per estensione a partire dal nucleo dei trasformativi, che presumibilmente rappresenta anche il filo rosso che congiunge questo tipo derivazionale al suo progenitore latino in *-tus*, come suggerisce la ricostruzione di FABRIZIO 2015 e FABRIZIO in stampa. Infatti, non solo i verbi trasformativi sono un nucleo di basi importanti per il suffisso latino (*adventus*, *exitus*, ecc.), ma sono centrali anche tra i primi tipi di femminili testimoniati in latino tardo (*ascensa*, *intrata*, ecc.).

Ancora più significativo è il dato che riguarda i derivati denominali, che in realtà riproducono la stessa *facies* semantica di operatori di *packaging*, stavolta dell'attività in cui è coinvolto il nome designato dalla base. A tal proposito, si registrano i seguenti cinque tipi semantici (cfr. GAETA 2002: 141; ACQUAVIVA 2005):

- | | | |
|-----|-----------------------------------|------------------------------|
| (6) | a. Colpo di N | <i>coltello / coltellata</i> |
| | b. Tipica azione di N | <i>spaccone / spacconata</i> |
| | c. Quantità contenuta in N | <i>cucchiaio / cucchiata</i> |
| | d. Intensificazione di N (spazio) | <i>valle / vallata</i> |
| | e. Intensificazione di N (tempo) | <i>notte / nottata</i> |

Gli ultimi due sono i meno produttivi, oltre che meno trasparenti dal punto di vista semantico. Li ho indicati per mezzo della perifrasi "Intensificazione di N" perché in effetti il nome in *-ata* fa riferimento ai due domini concettuali basici di spazio e di tempo e ne intensifica l'estensione per mezzo di un procedimento di individualizzazione. In altre parole, una *vallata* non è una semplice *valle*, ma una istanziazione di *valle* intesa come individualizzata in tutta la sua estensione, che quindi risulta intensificata. Lo stesso vale grosso modo per *nottata*. In questo modo si arriva a un valore che qualcuno chiama "collettivo" nel caso dello spazio e a un valore identificato spesso come "durativo" in quello del tempo¹¹. Riassumendo, il significato complessivo del pattern derivazionale può essere compreso dalla peri-

¹¹ Inoltre, il valore di intensificazione corrisponde a *pattern* molto **produttivi** in spagnolo, lingua in cui i derivati da FPP sono molto **produttivi**, analogamente e probabilmente più – specie nelle varietà sudamericane – che in italiano (cfr. GAETA 2015).

frasi: ‘singola istanziazione di N’, con focus sulle proprietà tipicamente associate a N. Si noti che in diversi casi la scelta della proprietà è determinata contestualmente: per esempio, *padellata* può valere sia come ‘colpo di N’ che ‘quantità contenuta in N’, mentre *bicchierata* in genere seleziona solo questo secondo valore. Vediamo ora più nel dettaglio quale risultato ci offre MIDIA (Fig. 8):

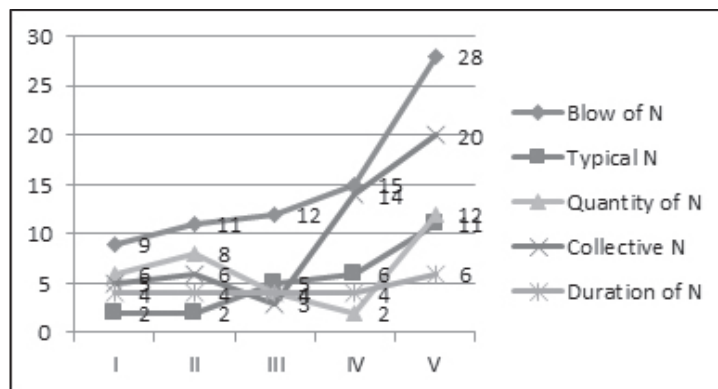


Fig. 8: I denominali in -ata in MIDIA.

Come si vede dal grafico, lo sviluppo diacronico è estremamente interessante, soprattutto alla luce del fatto che nella fase più antica il *pattern* è scarsamente rappresentato in confronto al computo complessivo dei FPP visto in Fig. 7. Inoltre, nella fase più antica i vari tipi sono per lo più sullo stesso piano, con il tipo ‘Colpo di N’ (= *Blow of N*) in testa e il tipo ‘Azione tipica di N’ (= *Typical N*) in fondo e gli altri ‘Quantità contenuta in N’ (= *Quantity of N*), ‘Intensificazione di N (spazio)’ (= *Collective N*) e ‘Intensificazione di N (tempo)’ (= *Duration of N*) in ordine decrescente. Il quadro cambia sostanzialmente arrivando al periodo più recente, con un incremento molto significativo del tipo già in testa ‘Colpo di N’, ma anche degli altri due ‘Azione tipica di N’ e ‘Quantità contenuta in N’. Si noti che anche il tipo ‘Intensificazione di N (spazio)’ risulta fortemente incrementato, a differenza di quello ‘Intensificazione di N (tempo)’, cosa piuttosto sorprendente rispetto alle valutazioni fatte in genere riguardo all’italiano attuale, che lo considerano scarsamente produttivo (cfr. RAINER 2004b: 254). Ma valgono a questo punto di nuovo le considerazioni fatte sopra a proposito degli altri nomi deverbali: le indicazioni fornite da MIDIA sono preziose, ma richiedono ulteriori indagini mirate, auspicate per il futuro.

5. Conclusioni

MIDIA si qualifica come uno strumento importante per svolgere ricerche nel dominio della formazione delle parole. La ricerca basata su MIDIA è estremamente stimolante perché da un lato porta dati freschi anche in confronto a fonti lessicografiche di grandi dimensioni e dall'altro permette di catturare tendenze diacroniche significative nella misura in cui il corpus è bilanciato. Per quanto riguarda i nomi deverbali indagati in questo contributo, abbiamo visto come le direzioni di ricerca messe in evidenza grazie a MIDIA confermino analisi già svolte e forniscano suggerimenti preziosi per eventuali indagini future. I nomi d'agente mostrano una preferenza abbastanza chiara per la selezione del morfoma costituito dal tema verbale, oltre che alcune interessanti microtendenze evolutive in relazione al genere testuale. D'altro canto MIDIA ci permette di individuare chiaramente lo sviluppo dei deverbali FPP a partire da un nucleo primario di selezione di verbi trasformativi che si è poi esteso ai continuativi con funzione di *packaging operator* e allo stesso tempo alle basi nominali. Da questo punto di vista, i corpora bilanciati offrono prospettive nuove per lo studio e l'interpretazione dei fatti morfologici e più in generale della dinamica lessicale dal punto di vista diacronico, anche se chiaramente non possono sostituirsi a lavori più mirati ed esaustivi cui auspicabilmente MIDIA contribuirà a dar corso in futuro.

BIBLIOGRAFIA

- ACQUAVIVA 2005 = PAOLO ACQUAVIVA, *I significati delle nominalizzazioni in -ATA e i loro correlati morfologici*, in MARIA GROSSMANN / ANNA M. THORNTON (edd.), *La formazione delle parole. Atti del XXXVII Convegno della Società Linguistica Italiana* (L'Aquila, 25-27 settembre 2003), Roma, Bulzoni: 7-27.
- ARONOFF 1976 = MARK ARONOFF, *Word-formation in generative grammar*, Cambridge (MA), MIT Press.
- ARONOFF 1994 = MARK ARONOFF, *Morphology by itself*, Cambridge (MA), MIT Press.
- BAAYEN 1989 = R. HARALD BAAYEN, *A corpus-based approach to morphological productivity. Statistical analysis and psycholinguistic interpretation*, Ph.D. Diss., Vrije Universiteit Amsterdam.
- BAAYEN 1994 = R. HARALD BAAYEN, *Productivity in language production*, in «Language and Cognitive Processes» 9: 447-469.
- BAAYEN 2009 = R. HARALD BAAYEN, *Corpus linguistics in morphology. Morphological productivity*, in LÜDELING/KYTÖ 2008-2009: II, 899-919.
- BARONI 2010 = MARCO BARONI, s.v. *corpora di italiano*, in RAFFAELE SIMONE (ed.), *Enciclopedia dell'italiano*, Roma, Istituto della Enciclopedia Italiana: I, 300-303.

- BARONI *et al.* 2004 = MARCO BARONI / SILVIA BERNARDINI / FEDERICA COMASTRI / LORENZO PICCIONI / ALESSANDRA VOLPI / GUY ASTON / MARCO MAZZOLENI, *Introducing the la Repubblica corpus. A large, annotated, TEI(XML)-compliant corpus of newspaper Italian*, in MARIA TERESA LINO / MARIA FRANCISCA XAVIER / FÁTIMA FERREIRA / RUTE COSTA / RAQUEL SILVA (edd.), *Proceedings of the 4th International conference on language resources and evaluation LREC* (Lisboa, 26-28/5/2004), Paris, ELRA European Language Resources Association: 1771-1774.
- BARONI *et al.* 2009 = MARCO BARONI / SILVIA BERNARDINI / ADRIANO FERRARESI / EROS ZANCHETTA, *The WaCky Wide Web. A collection of very large linguistically processed web-crawled corpora*, in «Language resources and evaluation» 43/3: 209-231.
- BARTSCH 1986 = RENATE BARTSCH, *On aspectual properties of Dutch and German nominalizations*, in VINCENZO LO CASCIO / CO VET (edd.), *Temporal structure in sentence and discourse*, Dordrecht, Foris: 7-39.
- BAUER 2001 = LAURIE BAUER, *Morphological productivity*, Cambridge, Cambridge University Press.
- BERTINETTO 1986 = PIER MARCO BERTINETTO, *Tempo, aspetto e azione nel verbo italiano*, Firenze, Accademia della Crusca.
- BOOIJ 2010 = GEERT BOOIJ, *Construction morphology*, Oxford, Oxford University Press.
- CICA = JOAN TORRUELLA (ed.), *Corpus Informatitzat del Català Antic*, in rete all'indirizzo <http://www.cica.cat/>.
- CLARIDGE 2008 = CLAUDIA CLARIDGE, *Historical corpora*, in LÜDELING/KYTÖ 2008-2009: I, 242-259.
- CoLFIS = PIER MARCO BERTINETTO / CRISTINA BURANI / ALESSANDRO LAUDANNA / LUCIA MARCONI / DANIELA RATTI / CLAUDIA ROLANDO / ANNA MARIA THORNTON, *Corpus e Lessico di Frequenza dell'Italiano Scritto*, in rete all'indirizzo <http://www.istc.cnr.it/grouppage/colfis>.
- Corpus «la Repubblica»* = *Corpus «la Repubblica»*, in rete all'indirizzo <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>.
- Corpus «La Stampa»* = *Corpus «La Stampa»*, 1996-1998, Università di Torino.
- COWIE 2000 = CLAIRE COWIE, *The discourse motivations for neologising. Action nominalization in the history of English*, in JULIE COLEMAN / CHRISTIAN J. KAY (edd.), *Lexicology, semantics and lexicography*, Amsterdam / Philadelphia, John Benjamins: 179-208.
- COWIE/DALTON-PUFFER 2002 = CLAIRE COWIE / CHRISTIANE DALTON-PUFFER, *Diachronic word-formation. Theoretical and methodological considerations*, in JAVIER E. DÍAZ VERA (ed.), *A changing world of words. Studies in English historical semantics and lexis*, Amsterdam / New York, Rodopi: 410-436.
- FABRIZIO 2015 = CLAUDIA FABRIZIO, *On Italian -ata action nouns. A diachronic account*, comunicazione presentata al 22nd International congress on historical lin-

guistics (Napoli 27-31/7/2015).

- FABRIZIO in stampa = CLAUDIA FABRIZIO, *Sulla nascita dei nomi in -ata nella transizione latino-italiano*, in «Rendiconti dell'Accademia Nazionale dei Lincei».
- FRANTEXT = *Base textuelle FRANTEXT*, ATILF - CNRS / Université de Lorraine, accessibile in rete previo abbonamento all'indirizzo <http://www.frantext.fr>.
- GAETA 2000 = LIVIO GAETA, *On the interaction between morphology and semantics. The Italian suffix -ATA*, in «Acta Linguistica Hungarica» 47/1-4: 205-229.
- GAETA 2002 = LIVIO GAETA, *Quando i verbi compaiono come nomi. Un saggio di morfologia naturale*, Milano, FrancoAngeli.
- GAETA 2003 = LIVIO GAETA, *Per una morfologia cognitiva*, in LIVIO GAETA / SILVIA LURAGHI (edd.), *Introduzione alla linguistica cognitiva*, Roma, Carocci: 101-122.
- GAETA 2004 = LIVIO GAETA, *Nomi d'azione*, in GROSSMANN/RAINER 2004: 314-351.
- GAETA 2007 = LIVIO GAETA, *On the double nature of productivity in inflectional morphology*, in «Morphology» 17/2: 181-205.
- GAETA 2010 = LIVIO GAETA, *On the viability of cognitive morphology for explaining language change*, in ALEXANDER ONYSKO / SASCHA MICHEL (edd.), *Cognitive Perspectives on Word Formation*, Berlin / New York, de Gruyter Mouton: 75-95.
- GAETA 2012 = LIVIO GAETA, *Strumenti come agenti: a metà tra natura e storia*, in VALENTINA BAMBINI / IRENE RICCI / PIER MARCO BERTINETTO (edd.), *Linguaggio e cervello – Semantica / Language and the Brain – Semantics*. Atti del XLII Congresso Internazionale di Studi della Società di Linguistica Italiana (Pisa, SNS, 2008), Roma, Bulzoni, CD-ROM: II.A.1.
- GAETA 2015 = LIVIO GAETA, *Action nouns in Romance*, in MÜLLER *et al.* 2015-2016: II, 1165-1185.
- GAETA/RICCA 2002 = LIVIO GAETA / DAVIDE RICCA, *Corpora testuali e produttività morfologica. I nomi d'azione italiani in due annate della Stampa*, in ROLAND BAUER / HANS GOEBL (edd.), *Parallela IX. Testo variazione informatica / Text Variation Informatik*. Atti del IX Incontro italo-austriaco dei linguisti (Salisburgo, 1-4/11/2000), Wilhelmsfeld, Egert: 223-249.
- GAETA/RICCA 2003 = LIVIO GAETA / DAVIDE RICCA, *Frequency and productivity in Italian derivation. A comparison between corpus-based and lexicographical data*, in «Italian Journal of Linguistics / Rivista di Linguistica» 15/1: 63-98.
- GAETA/RICCA 2005 = LIVIO GAETA / DAVIDE RICCA, *Aspetti quantitativi della produttività morfologica*, in TULLIO DE MAURO / ISABELLA CHIARI (edd.), *Parole e numeri*, Roma, Aracne: 107-124.
- GAETA/RICCA 2006 = LIVIO GAETA / DAVIDE RICCA, *Productivity in Italian word-formation. A variable-corpus approach*, in «Linguistics» 44/1: 57-89.
- GAETA/RICCA 2015 = LIVIO GAETA / DAVIDE RICCA, *Productivity*, in MÜLLER *et al.* 2015-2016: II, 841-858.

- GDLI = SALVATORE BATTAGLIA, *Grande dizionario della lingua italiana*, Torino, UTET, 1961-2002, 21 voll. (con 2 supplementi, 2004 e 2009).
- GRADIT = TULLIO DE MAURO (ed.), *Grande dizionario italiano dell'uso*, Torino, UTET, 1999, 6 voll. (con 2 supplementi, voll. VII e VIII, *Nuove parole italiane dell'uso*, 2003 e 2007, con chiave USB).
- GROSSMANN/RAINER 2004 = MARIA GROSSMANN / FRANZ RAINER (edd.), *La formazione delle parole in italiano*, Tübingen, Niemeyer.
- HC = MATTI RISSANEN (ed.), *The Helsinki Corpus of English Texts*, in rete all'indirizzo <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/>.
- ItWac = *Italian Wacky (Web-As-Corpus Kool Yinitiative)*, accessibile in rete previa registrazione all'indirizzo <https://cqpweb.lancs.ac.uk/>.
- JACKENDOFF 1991 = RAY JACKENDOFF, *Parts and boundaries*, in «Cognition» 41: 9-45.
- LAUDANNA *et al.* 1995 = ALESSANDRO LAUDANNA / ANNA M. THORNTON / GIORGINA BROWN / CRISTINA BURANI / LUCIA MARCONI, *Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente*, in SERGIO BOLASCO / LUDOVIC LEBART / ANDRÉ SALEM (edd.), *JADT 1995. III Giornate internazionali di analisi statistica dei dati testuali* (Roma, 11-13/12/1995), Roma, Cisu: I, 103-109.
- LÜDELING/KYTÖ 2008-2009 = ANKE LÜDELING / MERJA KYTÖ (edd.), *Corpus linguistics. An international handbook*, Berlin / New York, de Gruyter Mouton, 2 voll.
- MD-P = MARK DAVIES / MICHAEL J. FERREIRA (edd.), *O corpus do português*, in rete all'indirizzo <http://www.corpusdoportugues.org/>.
- MD-S = MARK DAVIES (ed.), *Corpus del español*, in rete all'indirizzo <http://www.corpusdelespanol.org/>.
- MIDIA = MIDIA. *Morfologia dell'Italiano in Diacronia*, in rete all'indirizzo <http://www.corpusmidia.unito.it/>.
- MÜLLER *et al.* 2015-2016 = PETER O. MÜLLER / INGEBORG OHNHEISER / SUSAN OLSEN / FRANZ RAINER (edd.), *Word-formation. An international handbook of the languages of Europe*, Berlin, de Gruyter Mouton, 5 voll.
- OED-MD = MARK DAVIES (ed.), *Oxford English Dictionary*, in rete all'indirizzo <http://corpus.byu.edu/oed/>.
- OVI = Istituto Opera del Vocabolario Italiano, *Corpus OVI dell'Italiano antico*, in rete all'indirizzo <http://gattoweb.ovi.cnr.it/>.
- PLAG 1999 = INGO PLAG, *Morphological productivity. Structural constraints in English derivation*, Berlin / New York, de Gruyter Mouton.
- PLAG *et al.* 1999 = INGO PLAG / CHRISTIANE DALTON-PUFFER / R. HARALD BAAYEN, *Productivity and register*, in «English Language and Linguistics» 3: 209-228.
- RAE = Banco de datos CORDE, *CORpus Diacrónico del Español*, Real Academia Española, in rete all'indirizzo <http://corpus.rae.es/cordenet.html>.
- RAINER 2004a = FRANZ RAINER, *L'origine dei nomi di strumento italiani in -tore*, in THOMAS KRISCH / THOMAS LINDNER / ULRICH MÜLLER (edd.), *Analecta homini universali dicata. Arbeiten zur Indogermanistik, Linguistik, Philologie, Poli-*

tik, Musik und Dichtung. Festschrift für Oswald Panagl zum 65. Geburtstag, Stuttgart, Heinz: I, 399-424.

RAINER 2004b = FRANZ RAINER, *Il suffisso -ata*, in GROSSMANN/RAINER 2004: 253-254.

SCALISE 1983 = SERGIO SCALISE, *Morfologia lessicale*, Padova, CLESP.

TALMY 1988 = LEONARD TALMY, *The relation of grammar to cognition*, in BRYGIDA RUDZKA-OSTYN (ed.), *Topics in cognitive linguistics*, Amsterdam / Philadelphia, John Benjamins: 165-205.

THORNTON 1990 = ANNA MARIA THORNTON, *Sui deverbali italiani in -mento e -zione (I)*, in «Archivio Glottologico Italiano» 75: 169-207.

THORNTON 1991 = ANNA MARIA THORNTON, *Sui deverbali italiani in -mento e -zione (II)*, in «Archivio Glottologico Italiano» 76: 79-102.

VENDLER 1967 = ZENO VENDLER, *Linguistics in philosophy*, Ithaca, Cornell University Press.

Zingarelli = Lo Zingarelli 1998. *Vocabolario della lingua italiana di Nicola Zingarelli*, a cura di MIRO DOGLIOTTI / LUIGI ROSIELLO, 12^a ed., Bologna, Zanichelli, 1997, con CD-ROM.

